Regression assumptions in clinical psychology research practice - A systematic review of common misconceptions

Anja Franziska Ernst [1] , **Casper J Albers** [Corresp. 1]

[1] Heymans Institute for Psychological Research, University of Groningen, Groningen, The Netherlands

Corresponding Author: Casper J Albers
Email address: c.j.albers@rug.nl

Misconceptions about the assumptions behind the standard linear regression model are widespread and dangerous. These lead to using linear regression when inappropriate, and to employing alternative procedures with less statistical power when unnecessary. Our systematic literature review investigated employment and reporting of assumption checks in twelve clinical psychology journals. Findings indicate that normality of the variables themselves, rather than of the errors, was wrongfully held for a necessary assumption in 4% of papers that use regression. Furthermore, 92% of all papers using linear regression were unclear about their assumption checks, violating APA-recommendations. This paper appeals for a heightened awareness for and increased transparency in the reporting of statistical assumption checking.

Regression Assumptions in Clinical Psychology Research Practice – A systematic review of

common misconceptions

Anja F. Ernst and Casper J. Albers[1]

University of Groningen

Heymans Institute for Psychological Research, Groningen, The Netherlands. [1] Corresponding

author: c.j.albers@rug.nl

3   Regression Assumptions in Clinical Psychology Research Practice – A systematic review of

4   common misconceptions

6   Abstract

7   Misconceptions about the assumptions behind the standard linear regression model are

8   widespread and dangerous. These lead to using linear regression when inappropriate, and to

9   employing alternative procedures with less statistical power when unnecessary. Our systematic

10   literature review investigated employment and reporting of assumption checks in twelve clinical

11   psychology journals. Findings indicate that normality of the variables themselves, rather than of

12   the errors, was wrongfully held for a necessary assumption in 4% of papers that use regression.

13   Furthermore, 92% of all papers using linear regression were unclear about their assumption

14   checks, violating APA-recommendations. This paper appeals for a heightened awareness for and

15   increased transparency in the reporting of statistical assumption checking.

18   Regression Assumptions in Research Practice – A systematic review of common misconceptions

20   One of the most frequently employed models to express the influence of several

21   predictors on a continuous outcome variable is the linear regression model:

22          $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i.$

23   This equation predicts the value of a case Y_i with values X_{ji} on the independent variables X_j ($j =$

24   $1, \ldots, p$). The standard regression model takes X_j to be measured without error (cf. Montgomery,

25   Peck & Vining, 2012, p.71). The various β_j slopes are each a measure of association between the

26   respective independent variable X_j and the dependent variable Y. The error for the given Y_i, the

27   difference between the observed value and value predicted by the population regression model, is

28   denoted by ε_i and is supposed to be unrelated to the values of X_p. Here, β_0 denotes the intercept,

29   the expected Y value when all predictors are equal to zero. The model includes p predictor

30   variables. In case $p = 1$, the model is denoted as the simple linear regression model.

31   The standard linear regression model is based on four assumptions. These postulate the

32   properties that the variables should have in the population. The regression model only provides

33   proper inference if the assumptions hold true (although the model is robust to mild violations of

34   these assumptions). Many statistical textbooks (for instance, Miles & Shevlin, 2001; Cohen,

35   Cohen, West & Aiken, 2003; Lomax & Hahs-Vaughn, 2012; Montgomery et al., 2012) provide

36   more background on these assumptions as well as advice on what to do when these assumptions

37   are violated.

38   Violations of these assumptions can lead to various types of problematic situations. First,

39   estimates may become biased, that is not estimating the true value on average. Second,

40   estimators may become inconsistent, implying that convergence to the true value when the

41   sample size increases is not guaranteed. Third, the ordinary least squares estimator may not be

42   efficient anymore: For instance, in the presence of assumption violations, OLS may provide less

43   accurate parameter estimates than other available estimation procedures. Fourth and finally,

44   NHST's and confidence intervals might become untrustworthy: p-values can be systematically

45   too small or too large, and confidence intervals are too narrow or too wide. This can occur even

46   if estimators are unbiased, consistent and efficient. For a more detailed description of these

47   issues, see Williams et al. (2013). Please note that these assumptions are the assumptions when

48   estimating using the Ordinary Least Squares (OLS) procedure, which is the default procedure in

49   many software packages, including SPSS and R. Other type of estimation methods, such as GLS,

50   apply other sets of assumptions.

51   Below, the four OLS-assumptions will be discussed.

52   ***Linearity.*** The conditional mean of the errors is assumed to be zero for any given

53   combination of values of the predictor variables. This implies that, for standard multiple

54   regression models, the relationship between every independent variable X_i and the population

55   mean of the dependent variable Y, denoted by μ_Y, is assumed to be linear when the other

56   variables are held constant. Furthermore, the relations between the various X_i and μ_Y are additive:

57   thus, the relation of X_i with μ_Y is the same, regardless of the value of X_j ($j \neq i$). This relates to the

58   issue of multicollinearity; a good model is expected to have as little overlap between predictors

59   as possible. However, multicollinearity is not a model assumption but merely a necessity for a

60   model to be parsimonious. Violation of this assumption can obviously occur when non-linear

61   relations are unmodelled, but also in case of measurement error (see Williams et al., 2013).

62   ***Normality.*** All errors are normally distributed around zero.

63   ***Homoscedasticity.*** The variance of the errors is the same for any combination of values

64   of the independent variables. Thus, this variance, which can then be denoted by a single symbol

65   (e.g. σ^2). This assumption is also called the homoscedasticity assumption. Thus, the second and

66   third regression assumptions combined specify that the errors (ε_i) of the model should follow a

67   normal distribution with a mean of zero and a (fixed) standard deviation σ. Heteroscedasticity

68   often manifests itself through a larger spread of measurements around the regression line at one

69   side of the scatterplot than at the other.

70   ***Independence.*** The errors $\varepsilon_1, \varepsilon_2, \ldots$, should be independent of one another: the pairwise

71   covariances should be zero. This assumption is not directly based on the distribution of the data

72   but on the study design and it requires the sampling method to be truly random (see, for instance,

73   Cohen, Cohen, West and Aiken, 2003). As with the normality assumption, scatterplots alone are

74   usually unsuitable for checking this assumption. A residual plot, or inspection of the

75   autocorrelation of the residuals, is a better approach.

76   ***Common misconceptions about assumptions.*** There are many misconceptions about the

77   regression model, most of which concern the assumptions of normality and homoscedasticity.

78   Most commonly, researchers incorrectly assume that X_i, or both X_i and Y, should be normally

79   distributed, rather than the errors of the model. This mistake was even made in a widely-read

80   article by Osborne and Waters (2002)attempting to educate about regression assumptions[1](cf.

81   Williams, Grajales & Kurkiewics, 2013).

82   Not assuming a normal distribution for X_i may seem counterintuitive at first, however the

83   indulgence of this assumption becomes more evident with an illustrative example. Take the

84   standard Student's t-test which assesses if two distributions are statistically different from one

---

[1]   Based on the journal's access counter, there were more than 540,000 views at the time of this writing
http://pareonline.net/genpare.asp?wh=0&abt=8

85   another (e.g., a t-test that compares the efficacy of a specific treatment compared to a placebo

86   treatment). The population distributions in both conditions are assumed to be normally

87   distributed with equal variances. This t-test can also be expressed as a regression model where

88   the independent variable X dummy codes the group membership, (i.e. if a participant is in the

89   control = 0, or in the treatment condition, $X = 1$). This regression model and the t-test are

90   mathematically equivalent and will thus lead to identical inference. Variable X will only attain

91   two values, 0 and 1, as it is only used as label for group membership. The dependent variable Y

92   will attain many different values: following a normal distribution for the treatment group and a

93   (possibly other) normal distribution for the control group. This resulting 'condition membership'

94   distribution is nothing close to normal (as it takes on just two values), however no assumption of

95   the general linear model is violated because the *subpopulations* of Y for each of the X values

96   follow a normal distribution with equal variances, as is visualised in Figure 1. This example

97   demonstrates that the assumptions of the t-test (standard normal distribution of the populations

98   around the group mean and equal variances) coincide with the second regression assumption.

99

100   As a consequence of the second regression assumption, the distribution of the dependent

101   variable conditional on some combination of values of the predictor variables is linear. Thus, Y_i

102   is actually normally distributed around μ_Y, the true conditional population mean. This becomes

103   clear when remembering that the error of the regression estimation is normally distributed

104   around mean zero and that Y_i is equal to $\mu_Y + \varepsilon_i$. That is, individual observations are the sum of

105   the mean and a deviation from this mean. However, it is wrong to test the normality of the

106   marginal distribution of the dependent variable Y because this would imply that all μ_Y values are

107   the same which is, generally, not the case. (This situation occurs only when all regression slopes

108   are zero and, thus, all predictor variables are linearly unrelated to Y.)

109   Regarding the linearity assumption, a common misconception is in thinking that only

110   linear relationships can be modelled using the OLS framework. This is not the case: the linearity

111   assumption deals with linearity in the parameters and the estimates, but not necessarily in the

112   variables.

113   ***Consequences of violations of assumptions*.** Misconceptions like the ones outlined above

114   potentially have severe effects on the ability to draw inferences from a data analysis. First of all,

115   the checking of wrong assumptions will most likely lead to the neglect of correct assumption

116   checking. If the researcher will decide on a regression analysis without having tested the correct

117   assumptions it is possible that some requirements of linear regression were not met. However, in

118   any case the neglect of correct assumption checking will always leave the reader or reviewer

119   unable to trust the results because there is no way of knowing whether the model assumptions

120   were actually met. Of course, the severity of this problem of non-transparency persists even

121   when the researcher ensured the validity of all necessary assumptions and merely failed to report

122   those findings. Not only does such non-transparency in data analysis lead to confusion for

123   researchers that are potentially interested in replicating or comparing the results, it also weakens

124   the informational value of the research findings that are being interpreted.

125   A second problem that is caused by misconceptions about model assumptions occurs

126   when a researcher decides against a linear regression analysis because of the violation of faulty

127   assumptions that were unnecessary in the first place. The difficulty of abandoning linear

128   regression analysis for a non-parametric procedure is the fact that the ordinary least squares

129   method of linear regression is a more powerful procedure than any of its non-parametric

130   counterparts, if its assumptions are met. Hence, wrongfully deciding against the employment of

131   linear regression in a data analysis will lead to a decrease in power. Thus, the understanding of

132   the correct regression assumptions is crucial because it prevents the abandonment of the linear

133   regression technique in cases in which it would be unjustified. Furthermore, the checking of

134   assumptions has another advantage: it might help the researcher to think about conceptually

135   alternative models. For instance, heteroscedasticity in the data could be a sign of an interaction

136   between one or more of the included independent variables and an independent variable not (yet)

137   included in the model.

138   Applying a linear regression model when assumptions are violated can lead to (severe)

139   problems, but this does not have to be the case, depending on the type of violation. Violations of

140   the linearity assumption and of the independence assumption can lead to biased, inconsistent and

141   inefficient estimates (Chatterjee & Hadi, 2006; Williams et al., 2013). A proper check on these

142   two assumptions is thus vital. The consequences of violations are less severe for the other two

143   assumptions.

144   If normality of errors holds, the OLS method is the most efficient unbiased estimation

145   procedure (White & MacDonald, 1980). If this assumption does not hold (but the remaining

146   assumptions do), OLS is only most efficient in the class of linear estimators (see Williams et al.,

147   2013, for a detailed discussion). This implies that, as long as the other assumptions are met,

148   estimates will still be unbiased and consistent in the presence of a normality violation, but the p-

149   values might be biased. Furthermore, the central limit theorem implies that for large samples the

150   sampling distribution of the parameters will be at least approximately normal, even if the

151   distribution of the errors is not. Hence, the regression model is robust with respect to violations

152   of the normality assumption. Potential problems will, in practice, primarily occur for inferential

153   problems (such as confidence intervals and testing) with small samples.

154   Similarly, violations of the homoscedasticity assumption are not necessarily problematic.

155   Provided that the very mild assumption of finite variance holds, estimates will still be unbiased

156   and consistent (Chatterjee & Hadi, 2006).

157   ***Best practices for checking of assumptions.*** There are many different ways to check the

158   four assumptions of the regression model and there generally is no 'uniformly optimal' approach.

159   Generally, there are two classes of approaches: (i) formal tests (of the style 'H_0: the

160   assumption is true' vs 'H_A: the assumption is violated') and (ii) graphical methods. For the

161   normality assumption alone, there is an abundance of formal tests, such as the Shapiro-Wilk test,

162   the Anderson-Darling test and the Kolmogorov-Smirnov test. Which approach is most powerful

163   depends on the kind of violation from normality (Razali & Wah, 2011). However, the use of

164   formal tests is discouraged by some (Albers, Boon & Kallenberg, 2000, 2001). When the

165   normality assumption holds, the null hypothesis of normality will still be rejected in α (usually

166   5%) of cases. This distorts the *p*-value distribution of the estimates of the regression model, even

167   when no assumptions are violated. Furthermore, tests for normality only have adequate power in

168   case of large sample sizes. However, when the sample size is large, the central limit theorem

169   implies that violations of normality have only limited effect on the accuracy of the estimates.

170   Applying graphical methods is therefore a preferred approach. This is also suggested by

171   the statistical guidelines for the APA set up by Wilkinson et al. (1999, p. 598): "Do not use

172   distributional tests and statistical indices of shape (e.g. skewness, kurtosis) as a substitute for

173   examining your residuals graphically". This advice builds upon the adagium by Chambers et al.

174   (1983, p. 1) that "there is no single statistical tool that is as powerful as a well-chosen graph". A

175 graph simply provides more information on an assumption than a single *p*-value ever can (see

176 also Chatterjee & Hadi, 2006, Ch. 4).

177   The linearity assumption can easily be checked using scatterplots or residual plots: plots

178 of the residuals vs. either the predicted values of the dependent variable or against (one of) the

179 independent variable(s). Note that residuals are the differences between the observed values and

180 the values predicted by the *sample* regression model, whereas errors denote the difference with

181 the values predicted by the *population* regression model. Residual plots are also the best visual

182 check for homoscedasticity. For the normality assumption, it is difficult to judge on the basis of a

183 scatterplot whether the assumption is violated. A histogram of the residuals is also a poor visual

184 check, as the 'shape' of the histogram heavily depends on the arbitrary choice of the bin width,

185 especially in small samples. Normal probability plots, or QQ-plots, provide a much better way to

186 check normality. Finally, a check on the independence assumption is done by studying the

187 autocorrelation function of the residuals. Note that this latter check does check for temporal

188 dependence violations of the independence assumptions, but not for other possible violations

189 such as clustering of observations. Furthermore, a common violation of independence involves

190 repeated-measures designs in which each individual contributes a set of correlated responses to

191 the data because of individual differences.

192

193   ***Outline of this paper.*** Misconceptions about frequently employed statistical tools, like

194 the *p*-value, are not rare, even amongst researchers (see Bakker and Wicherts, 2011; Hoekstra,

195 Morey, Rouder & Wagenmakers, 2014). Our paper aims to shed light onto potential

196 misconceptions researchers and reviewers might hold about the linear regression model.

197 Therefore, the documentary practices of psychological research papers with the linear regression

198   model and its assumptions were investigated by means of a literature review. In this review, we

199   investigate the proportion of papers where misconceptions around the assumptions of the

200   statistical regression model occurred and which type of misconceptions occurred most often.

201   This will provide important information, as the first step in solving flawed methodology in

202   research is finding out where the flaws are and how predominant they are.

203   Although the consequences of incorrectly dealing with assumptions can be severe, the

204   APA manual (American Psychological Association, 2010) barely provides guidelines on what to

205   report and how to report. It does *recommend* being specific about "information concerning

206   problems with statistical assumptions and/or data distributions that could affect the validity of

207   findings" (p. 248) as part of the Journal Article Reporting Standards, but this is not obligatory.

208   The APA Task Force on Statistical Inference (Wilkinson and Task Force on Statistical Inference,

209   1999) is more explicit in their recommendations: "You should take efforts to assure that the

210   underlying assumptions required for the analysis are reasonable given the data. Examine the

211   residuals carefully." (p. 598).

212   In this manuscript we present the findings of our literature review. Because the whole field of

213   psychological science is too broad to study in a single paper, we restrict ourselves to the field of

214   clinical psychology. We investigate how statistical assumptions were covered in various journals

215   of clinical psychology and what types of misconceptions and mistakes are occurring most often.

216   In the discussion section, possible explanations for the reported findings will be offered. The

217   paper will conclude with several proposals of how potential shortcomings in the current practices

218   with linear regression analysis could be overcome.

219   **Method**

220 **Journals.** The literature review restricted itself to articles that were published in clinical

221 psychology journals in the year 2013. It is possible that problems with the checking of

222 assumptions are less (or more) prominent in journals with a high impact, which is why we aimed

223 for a selection of journals with varied impact factors. We employed the Scientific Journal

224 Rankings (SJR) as reported on 16 December 2014 by the SCImago Journal and Country Rank

225 (SCImago, 2007) for clinical psychology journals of the year 2013 to divide all clinical

226 psychology journals into four quartiles (Q1 – Q4), where Q1 contains the 25% of journals with

227 the highest journal rank, etcetera. From every quartile the three highest ranked journals were

228 selected to be included in the review. Hence, we obtained a balanced selection from all clinical

229 psychology journals, as listed in Table 1. All articles published in the selected journals in 2013

230 were included, including also papers that had potentially been published online earlier. Letters,

231 journal corrigenda, editorial board articles and book reviews were not included in the review.

232 Basically, articles that were by design not containing a method section were not included in our

233 lists of articles. The focus of this review purely lies on published scientific articles.

234   Every article was retrieved directly from the official website of its respective journal

235 (except for Q1.3 which was directly retrieved from its official database "PsycARTICLES"). All

236 articles were in German (Q3.1), Spanish (part of Q3.3) or in English (all other). German articles

237 were also included in the review; Spanish articles were excluded because of the authors' lack of

238 proficiency in this language. Figure 2 displays the Prisma workflow of the analysis. We

239 conducted our review adhering to common meta-regression guidelines (Moher, Liberati,

240 Tetzlaff, Altman, The PRISMA Group, 2009).

241 **Procedure.**We evaluated whether and how papers described careful examination of the data with

242 regard to the underlying model assumptions whenever conducting statistical analysis (APA,

243   2010; Wilkinson et al., 1999). Papers were skimmed for the following criteria: if they had used

244   linear regression, how they tested the regression assumptions or what kind of assumptions they

245   indicated as being necessary, if they had transformed data on basis of correct or incorrect

246   assumptions and if a paper had considered an ordinary least squares regression model but

247   employed a different model on basis of either correct or incorrect assumptions. This resulted in a

248   classification scheme of 12 different rubrics which are displayed in Table 2. This scheme is

249   mutually exclusive and exhaustive; all studied papers are classified into exactly one rubric. The

250   search strategy has been carried out by Anja Ernst. Independently, Casper Albers checked and

251   classified 10% of the manuscripts in the Q1-journals. No mismatch between both sets of

252   classifications occurred.

253   Papers that used linear regression were classified as follows. We assumed the most

254   common misconception about linear regression to be the checking of the normality of the

255   variables while failing to check the normality of the errors. Therefore, we created rubrics 8 to 11

256   to classify all papers that employed linear regression and checked or assumed the normality of X

257   and/or Y but not of the errors. An example of a paper classified in rubric 8 stated "Variable

258   distributions were tested to ensure assumptions of normality, linearity, and variance equality

259   were met, with no significant violations observed" (Nadeau, Lewin, Arnold, Crawford, Murphy

260   & Storch, 2013). Often, when the normality assumption was mentioned it was unclear whether

261   authors had checked the normality of errors or of the variables. Articles that were unclear in this

262   regard were classified under rubric 5. For instance, one of the articles classified in this rubric

263   stated "Preliminary analysis examined data for the presence of outliers and the appropriateness

264   of assumptions of normality, linearity, and homoscedasticity" (Nguyen, Barrash, Koenigs,

265   Bechara, Tranel & Denburg, 2013) with no more information provided on the assumption

266   checks. Papers that indicated to have checked the homoscedasticity, normality of the errors and

267   linearity assumptions were classified as *'Correct'* in rubric 4. Articles that mentioned at least a

268   few correct assumptions, as opposed to giving no indication at all (rubric 7), were classified in

269   rubric 6. Because all papers that checked or assumed the normality of X or Y but not of the errors

270   were included in rubrics 8 to 11, we have named rubric 6 *'Did not test all but some correct*

271   *assumptions, did not include normality of variables'*. After performing the literature review it

272   became apparent that none of the articles listed in this category had mentioned the normality of

273   errors. Because we aimed to demonstrate how rare it is to read that researchers check the

274   normality of the errors we have updated the name of the category into '*Did not test all but some*

275   *correct assumptions, included neither normality of variables nor errors'*, even though the

276   checking of the normality of errors was not employed as a criterion for inclusion in this category

277   during the literature review.

278   Papers that did not fit into any of the eleven other rubrics but included an aspect on linear

279   regression assumptions that we found unsatisfactory were listed in the rubric *'Other*

280   *misconceptions about assumptions'*. One example of a paper classified in this category claimed

281   "All assumptions of multiple regression (linearity, multicollinearity, and homoscedasticity) were

282   met" this paper was included in the category *'Other misconceptions'* because they did not only

283   lack any mention whether normality of the residuals was checked (which would have resulted in

284   a classification in rubric 6) but also claimed that a list not containing normality of residuals was

285   complete. We found this claim unsatisfactory which was the reason we included this paper in

286   rubric 12.

287   Whenever an article in our selection reported the results of a regression analysis of

288   another paper or reviewed several linear regression articles, it was evaluated whether the paper

289   reviewing all the previous regression analysis had made it a criterion of inclusion whether the

290   assumptions have been met in the original articles. If a review article did not check or mention

291   the assumptions of the papers that published the original analysis, the article was classified as

292   '*Use of linear regression but no indication if any or which assumptions were tested*'. However,

293   these sorts of papers constitute less than one percent of our selected articles. It should be noted

294   that this only applies to papers which reported the data values of a linear regression or analysed

295   regression results from other studies. A paper was not included if it only mentioned the direction

296   of the outcomes of another paper's regression model or stated that a relationship had been

297   established by previous research findings.

298   Because the focus of this paper lies on the assumptions of linear regression, only linear

299   regression model assumptions were examined in the literature review. Consequently, papers that

300   analysed data by means of other types of regression, such as latent factor models, logistic

301   regression, and proportional hazards models (Cox regression), were not inspected for assumption

302   checking. When a paper used a regression model other than linear regression, and without

303   mentioning that linear regression was alternatively considered for data analysis it was classified

304   as '*No Model of Interest*'.

305 **Results**

306   The results of the systematic literature review are displayed in Tables 3, 4 and 5 which

307   display the number of occurrences of different classifications for the selected journals. In the

308   online supplementary material we indicate for all of the 893 individual papers studied into which

309   category they fall.

310   Table 3 shows the findings for all journals with the 12 different classification rubrics

311   summarized into seven different columns. The three columns entitled 'Dealing with

312   assumptions' list the number of different types of regression papers in a specific journal and

313   shows the proportional amount of this type in relation to the complete number of regression

314   articles in that journal. The two columns for 'No regression' list the number of papers which did

315   not use a linear regression model and included in their method sections to have considered a

316   linear regression analysis but decided against it on the basis of checking either correct or

317   incorrect assumptions.

318   Table 4 specifies the details behind the articles which are listed in Table 3 under the

319   column titled 'incorrectly '. This table classifies the corresponding 10 papers into Rubrics 8 – 12

320   of Table 2. It may be noted that 4% of all articles that used linear regression checked normal

321   distributions of some variables instead of normal distribution of errors.

322   Table 5 specifies the details behind the column 'unclear' in Table 2; that is it classifies

323   the 159 corresponding papers into Rubrics 5 to 7 of Table 2. Of all papers that employed

324   regression, 92% were unclear about the assumptions of the linear regression model that were

325   tested or were thought to be fulfilled.

Discussion

326

327   In our analysis, we studied 893 papers, representative for the work published in the field

328   of clinical psychology, and classified the 172 papers (19.4%) which considered linear regression

329   into three categories: those that dealt with the assumptions correctly, those that dealt with

330   assumptions incorrectly, and those that did not specify how they dealt with assumptions.

331   Merely 2% of these papers were both transparent and correct in their dealing with

332   statistical assumptions. Furthermore, in 6% of papers transparency was given but the dealing

333   with assumptions was incorrect. Hoekstra, Kiers & Johnson (2012) might provide some insight

334   into why researchers did not check assumptions. They list unfamiliarity with either the fact that

335   the model rests on the assumption or with how to check the assumption as the top two reasons.

336   As explained, incorrect dealing with the assumptions could lead to severe problems regarding the

337   validity and power of the results. We hope that this manuscript creates new awareness of this

338   issue with editors of clinical psychology journals and that this will assist in bringing down the

339   number of publications with flawed statistical analyses.

340   A tremendous amount of papers that employed regression, 92% of those studied, were

341   not clear on how they dealt with assumptions. It is not possible (not for us, nor for the reader) to

342   judge from the text whether checks for assumption violations were performed correctly. In the

343   group of transparent papers, the number of papers with fundamental mistakes in dealing with

344   assumptions far outnumber the number of papers without mistakes. Thus, even though it is not

345   possible to pinpoint an exact number to it, it would be naive to assume that only a small

346   proportion of those 92% also deal with assumptions incorrectly.

347   We believe that most contemporary problems in the handling of regression methods

348   could be counteracted by a more thorough coverage of the statistical assumption checks that

349   were performed in order to determine the validity of the linear regression model. At the very

350   least, transparency regarding how assumptions are approached, in line with the recommendations

351   by Wilkinson et al. (1999), is essential. Thus, mentioning which assumptions were checked and

352   what diagnostic tools were used to check them under what criteria, should be a minimum

353   requirement. Preferably, the authors should also show the results of these checks.

354   With transparency, the critical reader can distinguish correct approaches from incorrect

355   ones, even if the author(s), editor(s) and referees fail to spot the flaws. These statistical checks

356   can be given in the paper itself, but could also be provided in online supplementary material, a

357   possibility most journals offer nowadays (note that none of the papers investigated in this

358   manuscript referred to supplementary material for assumption checks). Thus, increased length of

359   the manuscript does not need to be an issue. Our aspiration for an increased transparency in

360   statistical assumption checks is in line with recent developments in psychology such as open

361   methods (obligatory in e.g. the APA-journal Archives of Scientific Psychology) and open data

362   (either published as online supplementary material with a paper, or through special journals like

363   Journal of Open Psychology Data). With open data, sceptical scientists can re-do the analyses

364   and check assumptions for themselves. Enforcing, or at least strongly encouraging, transparency

365   can even have beneficial effects to the level of publications in the respective journal (Wicherts,

366   Bakker & Molenaar, 2011). Even if publishing the data does not have a direct beneficial effect

367   on the quality of work, it will be useful as it provides the sceptical reader with the required

368   information to perform the assumption checks and thus the possibility to check the credibility of

369   the published work.

370

371   It is difficult to establish whether high ranking journals deal with assumptions more

372   adequately than lower ranking journals. Even though the results in Table 5 indicate that higher

373   ranked journals were more likely to test at least a few assumptions compared to lower ranked

374   journals; the results do mainly show that there is great variability between journals regarding the

375   number of papers with applied regression models they publish: two journals published no papers

376   in 2013 that employed linear regression, and five journals published six or fewer of these papers.

377   Because two of the three inspected Q1 journals are review journals they predominantly

378   employed meta-regression, a special type of regression useful for conducting meta-analyses, and

379   only rarely linear regression, it should be pointed out that of the 15 papers that used meta-

380   regressions in our Q1.2 eleven tested at least some of the required assumptions (that is 73% of

381   meta-regression papers were checked correctly for statistical assumptions). We believe that for

382   these papers the percentage is much better than the overall percentage of 2% for applied

383   regression papers, because meta-analyses are usually carried out by a team of authors including

384   at least one statistician or psychometrician.

385   We have limited our literature review to papers employing linear regression models, in

386   order to keep the study feasible. We suspect that similar findings would arise when studying

387   other classes of statistical models. Furthermore, we have also limited the review to papers

388   published in the field of clinical psychology; however we suspect that similar problems occur –

389   albeit possibly in different proportions – in all areas of applied psychological research. Thus, our

390   suggestions with respect to increased transparency and better evaluation of the employed

391   methodology should be relevant for a wider range of papers than those studied here. Because our

392   categorization of papers is reasonably straightforward, only one author conducted most of the

393   review. While our rubrics allow objective classifications we cannot preclude a few single

394 accidental misclassifications. However, possible misclassification should be minimal at most and

395 can therefore be expected to not have skewed the overall results that are based on a large number

396 of papers. Thus, despite this limitation we are confident in the overall results. For future

397 research, it would be interesting to do a similar literature review based on either alternative

398 techniques or on another field of application. Furthermore, more research is needed in

399 understanding the reasons that underlie why researchers frequently do not check assumptions.

400 One of the consequences of the lack of reporting of assumption checks is that many

401 published findings in clinical psychology are underestimating the uncertainty in their claims. For

402 instance, reported confidence intervals in the literature describe the uncertainty surrounding the

403 parameter, if the OLS-assumptions are met. The uncertainty of the validity of the assumptions

404 should lead to wider confidence intervals, in general. For future research, it would be an

405 interesting puzzle to assess the magnitude of this added uncertainty.

406 To summarise, in order to prevent the observed problems that were outlined above we

407 suggest a more transparent methodological reporting. Research should cover which assumption

408 checks were carried out. Furthermore, it should be mentioned if alternative statistical models

409 have been considered and why they were not employed, if so. This will be a necessity for future

410 research articles in order to be able to detect and prevent errors related to widespread

411 misconceptions but also to remove doubt from articles with an actual immaculate data analysis.

412

413

414 **Additional information**

415 A detailed breakdown of the systematic review, references to all websites employed to retrieve

416 articles as well as a completed PRISMA checklist are provided as online supplementary material.

417

References

418

419

420   Albers, W., Boon, P. C., & Kallenberg, W. C. M. (2000). Size and power of pretest procedures.

421   *Annals of Statistics*, 28, 195-214. Retrieved from: http://www.jstor.org/stable/2673986.

422   Albers, W., Boon, P. C., & Kallenberg, W. C. M. (2001). Power gain by pre-testing? *Statistics &*

423   *Decisions*, 19(3), 254-276.

424   American Psychological Association (2010). *Publication Manual of the American*

425   *Psychological Association* (Sixth Edition). Washington D.C.: American Psychological

426   Association.

427   Bakker, M. & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology

428   journals. *Behavior Research Methods*, 43, 666-678, doi:10.3758/s13428-011-0089-5

429   Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. A. (1983). *Graphical Methods for*

430   *Data Analysis.* Pacific Grove, CA: Wadsworth & Brooks/Cole

431   Cohen, J., Cohen, P., West, S.G., Aiken, L.S. (2003). *Applied Multiple Regression/Correlation*

432   *Analysis for the Behavioral Sciences* (Third Edition). New York, NY: Routledge.

433   Chatterjee, S. & Hadi, A. S. (2006). *Regression Analysis by Example, Fourth Edition.* Hoboken,

434   NJ:

435   John Wiley & Sons.

436   Hoekstra, R., Kiers, H. A. L., & Johnson, A. L. (2012). Are assumptions of well-known

437   statistical techniques checked, and why (not)? *Frontiers in Psychology*, 3, 137,

438   doi:10.3389/fpsych.2012.00137

439   Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust

440   misinterpretation of confidence intervals. *Psychonomic Bulletin & Review,* 21, 1157–

441   1164. doi:10.3758/s13423-013-0572-3/21.12/2014

442   Lomax, R.G., & Hahs-Vaughn, D. L. (2012). Statistical Concepts: A Second Course. New York,

443   NY: Routhledge.

444   Miles, J. & Shevlin, M. (2001). Applying Regression and Correlation: A Guide for Students and

445   Researchers. London, UK: Sage.

446   Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group (2009). *P*referred

447   *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement.

448   PLoS Medicine 6(6): e1000097. doi:10.1371/journal.pmed1000097/21.12/2014

449   Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression

450   Analysis (5th edition). New York: John Wiley & Sons.

451   Nadeau, J. M., Lewin, A. B., Arnold, E. B., Crawford, E. A., Murphy, T. K., & Storch, E. A.

452   (2013). Clinical correlates of functional impairment in children and adolescents with

453   obsessive–compulsive disorder. *Journal Of Obsessive-Compulsive And Related*

454   *Disorders*, 2(4), 432-436. doi:10.1016/j.jocrd.2013.10.002

455   Nguyen, C. M., Barrash, J., Koenigs, A. L., Bechara, A., Tranel, D., & Denburg, N. L. (2013).

456   Decision-making deficits in normal elderly persons associated with executive personality

457   disturbances. *International Psychogeriatrics*, 25 (11), 1811-1819.

458   doi:10.1017/S1041610213001270

459   Osborne, J. & Waters, E. (2002). Four assumptions of multiple regression that researchers should

460   always test. *Practical Assessment, Research & Evaluation*, 8(2), 1-9. Retrievedfrom:

461   http://PAREonline.net/getvn.asp?v=8&n=2/16.12/2014

462   Razali, N. & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov,

463   Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*,

464   2(1): 21-33.

465   SCImago. (2007). *SJR — SCImago Journal & Country Rank.* http://www.scimagojr.com.

466   Retrieved

467   on: 16.12.2014.

468   Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics, 6th edition. London, UK:

469   Pearson.

470   Wicherts, J. M., Bakker, M. & Molenaar, D. (2011). Willingness to share research data is related

471   to the strength of the evidence and the quality of reporting of statistical results. *PLOS*

472   *One*, 6(11), doi: 10.1371/journal.pone.0026828

473   White, H. & MacDonald, G. M. (1980). Some large-sample tests for nonnormality in the linear

474   regression model. *Journal of the American Statistical Association*, 75(369): 16-28,

475   doi:10.2307/1912934.

476   Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical Methods in Psychology

477   Journals: Guidelines and Explanations. *American Psychologist,* Vol. 54, No. 8, 594–604

478   Williams, M. N., Grajales, C. A . G. G., & Kurkiewicz, D. (2013). Assumptions of multiple

479   regression: Correcting two misconceptions. *Practical Assessment, Research &*

480   *Evaluation,* 18(11), 1-14.

481

482 **Tables**

Label	Journal
Q1.1	Annual Review of Clinical Psychology
Q1.2	Clinical Psychology Review
Q1.3	Journal of Consulting and Clinical Psychology
Q2.1	International Psychogeriatrics
Q2.2	Journal of Attention Disorders
Q2.3	American Journal of Drug and Alcohol Abuse
Q3.1	Zeitschrift fur Klinische Psychologie und Psychotherapie
Q3.2	Journal of Obsessive-Compulsive and Related Disorders
Q3.3	International Journal of Psychology and Psychological Therapy
Q4.1	Internet Journal of Mental Health
Q4.2	Indian Journal of Psychological Medicine
Q4.3	Behaviour Change

483 *Table 1*: Selection of Clinical Psychology Journals. The first column gives the ranking of the

484 journal, the first number denoting the quartile in which the journal falls, the second number the

485 rank of the journal within that quartile.

486

Class.	Reason
Papers without a linear regression model:	
1	No Model of Interest
2	Rejection of linear regression on basis of correct assumptions
3	Rejection of linear regression on basis of not meeting incorrect assumptions
Papers with a linear regression model:	
4	Correct linear regression
5	Mentioned all correct assumptions but not if the 'normality assumption' was tested on the residuals or on X or Y
6	Did not test all but some correct assumptions, included neither normality of variables nor errors
7	Use of linear regression but no indication if any or which assumptions were tested
8	Assumed/tested normally distributed X but not the normality of the errors
9	Assumed/tested normally distributed Y but not the normality of the errors
10	Assumed/tested normally distributed X and Y but not the normality of the errors
11	Assumed/tested normally distributed variables but did not indicate if X or Y or both and did not test the normality of the errors
12	Other misconceptions about assumptions

487

488 *Table 2*: Classification of the reviewed regression papers. Rubrics 3 and 5 – 12 represent papers

489 with imperfect handling of regression assumptions: in rubrics 5 – 7 it is unclear from whether

490 assumptions are correctly dealt with; in rubrics 8 – 12 the dealing with assumptions was

491 incorrect.

492

Journal	Number of papers (rub. 1–12)	Number of papers with regression (rub. 4–12)	Dealing with assumptions			No regression	
			Correctly (rub. 4)	Unclear (rub. 5–7)	Incorrectly (rub. 8–12)	Correct (violation of true assumption) (rub. 2)	Incorrect (violation of false assumption) (rub. 3)
Q1.1	33	0	0	0	0	0	0
Q1.2	86	6 (7%)	0	6 (100%)	0	0	0
Q1.3	98	26 (28%)	0	25 (100%)	0	3 (100%)	0
Q2.1	227	44 (19%)	3 (7%)	39 (89%)	2 (5%)	1 (100%)	0
Q2.2	199	52 (26%)	0	49 (94%)	3 (6%)	0	0
Q2.3	54	14 (26%)	0	14(100%)	0	0	0
Q3.1	23	5 (22%)	0	5 (100%)	0	1 (50%)	1 (50%)
Q3.2	59	21 (55%)	0	16 (71%)	5 (29%)	1 (100%)	0
Q3.3*	10*	2 (20%)*	0*	2 (100%)*	0*	0*	0*
Q4.1	2	1 (50%)	0	1 (100%)	0	0	0
Q4.2	82	0	0	0	0	0	0
Q4.3	20	2 (10%)	0	2 (100%)	0	0	0
Total	893	172 (19 %)	3 (2%)	159 (92%)	10 (6%)	6 (86%)	1 (14%)

493

494 *Table 3*: Proportion of various types of papers in our selected journals. Categorisations are mutually

495 exclusive and exhaustive. Journals are referred by the labels assigned in Table 1. "Rub." refers to the

496 rubrics in Table 2. The online supplementary material indicates which papers belong to each of the

497 numbers in this table.

498 * Papers in Spanish excluded

499
500

Journal	Articles with flawed linear regression model (rub. 8-12)	Tested normality of X but not of residuals (rub. 8)	Tested normality of Y but not of residuals (rub. 9)	Assuming normally distributed variables but did not indicate if X or Y or both (rub.10)	Tested normality of X and of Y but not of residuals (rub. 11)	Other misconceptions (rub. 12)
Q2.1	2	0	0		0	2 (100%)
Q2.2	3	2 (67%)	0	0	0	1 (33%)
Q3.2	5	4 (80%)	1 (20%)	0	0	0
Total	10	6 (60%)	1 (10%)	0	0	3 (30%)

501
502   *Table 4*: Breakdown of the types of mistakes that were observed. Only Journals with flawed

503   models are listed. Categorizations are mutually exclusive and exhaustive. Journals are referred

504   by the labels assigned in Table 1.

505

Journal	Papers in which handling of regression assumption was unclear *(rub. 5-7)*	Unclear		
		if the 'normality assumption' was tested on the residuals or on *X* or *Y (rub. 5)*	Did not test all but some correct assumptions *(rub. 6)*	no indication if any or which assumptions were tested *(rub. 7)*
Q1.2	6	0	2 (33%)	4 (67%)
Q1.3	26	0	0	25 (100%)
Q2.1	39	4 (10%)	5 (13%)	30 (77%)
Q2.2	49	1 (2%)	2 (4%)	46 (94%)
Q2.3	14	0	1 (7%)	13 (93%)
Q3.1	5	0	0	5 (100%)
Q3.2	16	0	0	16 (100%)
Q3.3	2	0	0	2 (100%)
Q4.1	1	0	0	1 (100%)
Q4.3	2	0	0	2(100%)
Total	159	5 (3%)	10 (6%)	144 (91%)

506

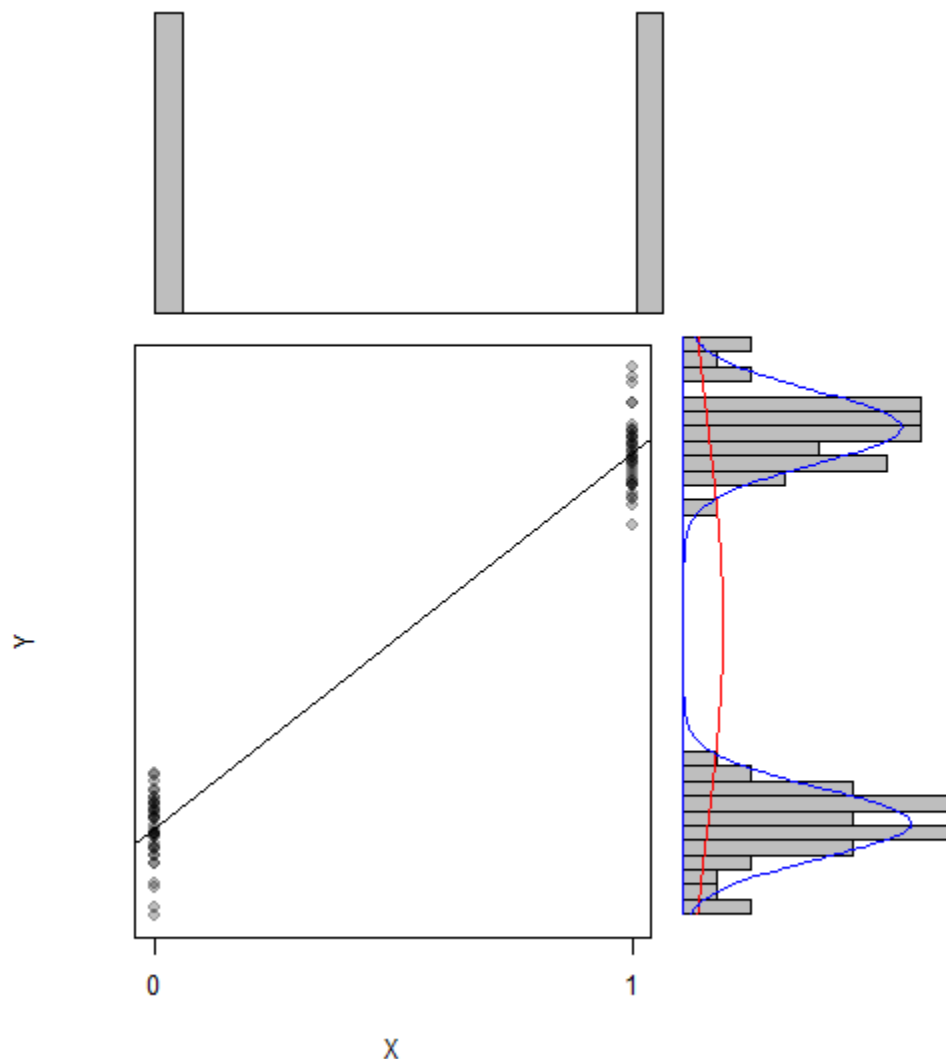
507   *Table 5*: Breakdown of the different types of 'Unclear' classifications. Only Journals with unclear

508   models are listed. Categorizations are mutually exclusive and exhaustive. Journals are referred by

509   the labels assigned in Table 1.

510

511 **Figures**

512

513 *Figure 1*: Simulated example of a *t*-test based on $n = 40$ observations per group and no violations

514 of the assumptions. The main panel shows a scatterplot of (X, Y)-scores. The red curve

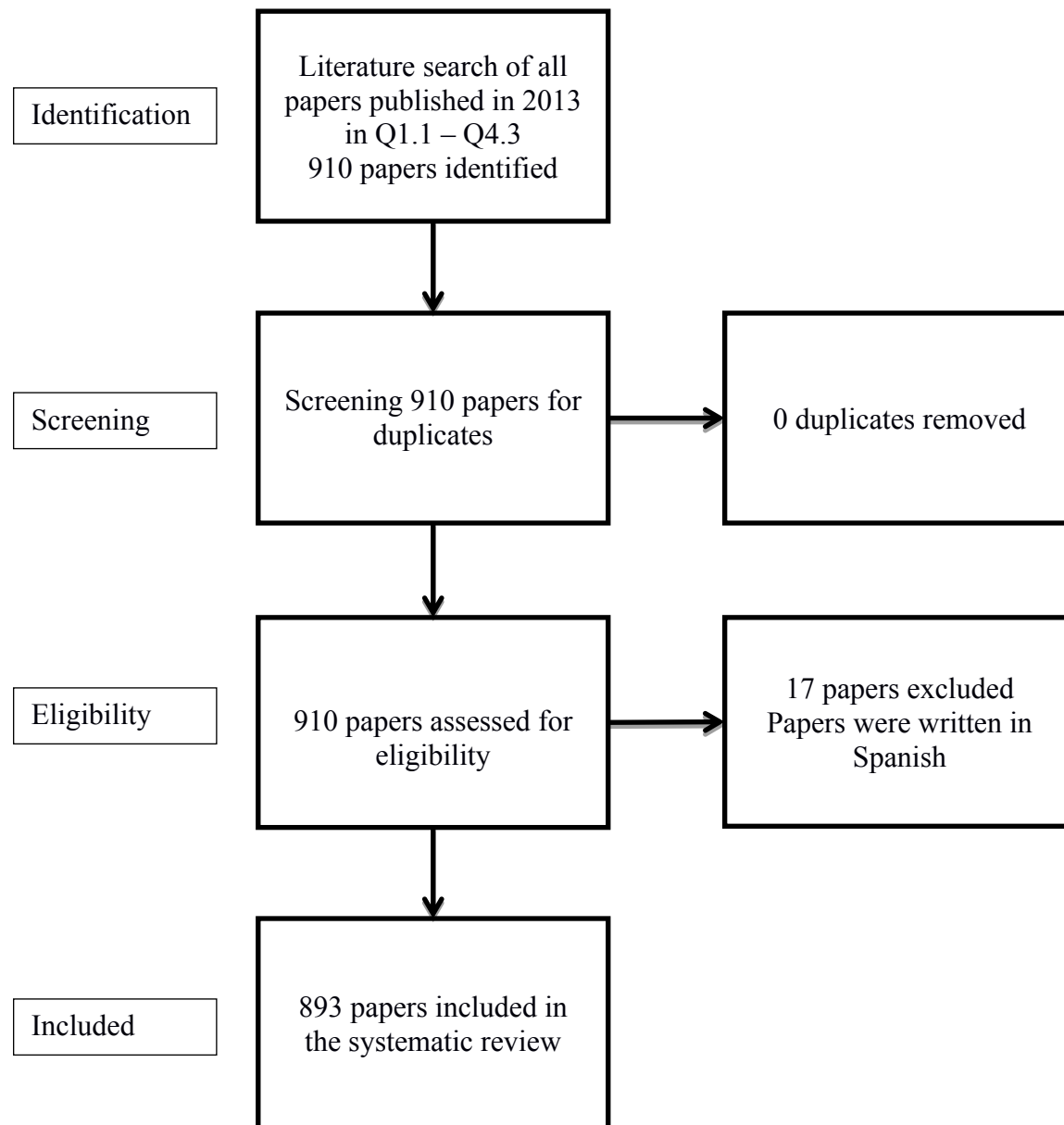
515 corresponds to the best-fitting normal distribution for Y, where the blue curves correspond to the

516 best-fitting normal distribution for both subpopulations of Y. The histograms in the top and side

517 panels clearly indicate non-normality for X and Y. However, within both subpopulations the

518 distribution is normal

519

Identification	Literature search of all papers published in 2013 in Q1.1 – Q4.3 910 papers identified	
Screening	Screening 910 papers for duplicates	0 duplicates removed
Eligibility	910 papers assessed for eligibility	17 papers excluded Papers were written in Spanish
Included	893 papers included in the systematic review	

520

521   *Figure 2*: Prisma flow diagram of included records